Эконометрика, 2020-2021, 3 модуль

Семинары 3-4

25.01.21, 1.02.21

Для Группы Э_Б2018_Э_3

Семинарист О.А.Демидова

Метод максимального правдоподобия
Проверка гипотез

Материал задачника Борзых Д. И Демешева Б.

Метод максимального правдоподобия

Пусть

$X = (X_1, \ldots, X_n)$ — случайная выборка

$x = (x_1, \ldots, x_n)$ — реализация данной случайной выборки

$f_{X_i}(x_i, \theta)$ — плотность распределения случайной величины X_i, $i = 1, \ldots, n$

$\theta = (\theta_1, \ldots, \theta_k)$ — вектор неизвестных параметров

$\Theta \subseteq \mathbb{R}^k$ — множество допустимых значений вектора неизвестных параметров

$\mathrm{L}(\theta) = \prod_{i=1}^{n} f_{X_i}(x_i, \theta)$ — функция правдоподобия

$\ell(\theta) := \ln \mathrm{L}(\theta)$ — логарифмическая функция правдоподобия

Пусть требуется протестировать систему (нелинейных) ограничений относительно вектора неизвестных параметров

$$H_0 : \begin{cases} g_1(\theta) = 0 \\ g_2(\theta) = 0 \\ \ldots \\ g_r(\theta) = 0 \end{cases}$$

где $g_i(\theta)$ — функция, которая задаёт i-ое ограничение на вектор параметров $\theta, i = 1, \ldots, r$.

$$\frac{\partial g}{\partial \theta'} = \begin{pmatrix} \partial g_1/\partial \theta' \\ \partial g_2/\partial \theta' \\ \vdots \\ \partial g_r/\partial \theta' \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_1}{\partial \theta_2} & \cdots & \frac{\partial g_1}{\partial \theta_k} \\ \frac{\partial g_2}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_2} & \cdots & \frac{\partial g_2}{\partial \theta_k} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_r}{\partial \theta_1} & \frac{\partial g_r}{\partial \theta_2} & \cdots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}$$

$$\frac{\partial g'}{\partial \theta} = \begin{pmatrix} \frac{\partial g_1'}{\partial \theta} & \frac{\partial g_2'}{\partial \theta} & \dots & \frac{\partial g_r'}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} & \frac{\partial g_2}{\partial \theta_1} & \dots & \frac{\partial g_r}{\partial \theta_1} \\ \frac{\partial g_1}{\partial \theta_2} & \frac{\partial g_2}{\partial \theta_2} & \dots & \frac{\partial g_r}{\partial \theta_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial g_1}{\partial \theta_k} & \frac{\partial g_2}{\partial \theta_k} & \dots & \frac{\partial g_r}{\partial \theta_k} \end{pmatrix}$$

$I(\theta) = -\mathbb{E}\left(\frac{\partial^2 \ell}{\partial \theta \partial \theta'}\right) = -\mathbb{E}\begin{pmatrix} \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_2 \partial \theta_k} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 \ell}{\partial \theta_k \partial \theta_k} \end{pmatrix}$ — информационная матрица Фишера

$$\frac{\partial \ell}{\partial \theta} = \begin{pmatrix} \frac{\partial \ell}{\partial \theta_1} \\ \frac{\partial \ell}{\partial \theta_2} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_k} \end{pmatrix}$$

$\Theta_{UR} := \Theta$ — множество допустимых значений вектора неизвестных параметров без учёта ограничений

$\Theta_R := \{\theta \in \Theta : g(\theta) = 0\}$ — множество допустимых значений вектора неизвестных параметров с учётом ограничений

$\hat{\theta}_{UR} \in \Theta_{UR}$ — точка максимума функции ℓ на множестве Θ_{UR}

$\hat{\theta}_R \in \Theta_R$ — точка максимума функции ℓ на множестве Θ_R

Тогда для тестирования гипотезы H_0 можно воспользоваться одной из следующих ниже статистик.

$LR := -2(\ell(\hat{\theta}_R) - \ell(\hat{\theta}_{UR})) \overset{a}{\sim} \chi^2_r$ — статистика отношения правдоподобия

$W := g'(\hat{\theta}_{UR}) \cdot \left[\frac{\partial g}{\partial \theta'}(\hat{\theta}_{UR}) \cdot I^{-1}(\hat{\theta}_{UR}) \cdot \frac{\partial g'}{\partial \theta}(\hat{\theta}_{UR})\right]^{-1} g(\hat{\theta}_{UR}) \overset{a}{\sim} \chi^2_r$ — статистика Вальда

$LM := \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R)\right]' \cdot I^{-1}(\hat{\theta}_R) \cdot \left[\frac{\partial \ell}{\partial \theta}(\hat{\theta}_R)\right] \overset{a}{\sim} \chi^2_r$ — статистика множителей Лагранжа

Задачи на нахождение оценок ММП

1) Из задачника Демешев-Борзых

5.4 Совместное распределение величин X и Y задано функцией

$$f(x, y) = \frac{\theta(\beta y)^x e^{-(\theta+\beta)y}}{x!}.$$

Величина X принимает целые неотрицательные значения, а величина Y — действительные неотрицательные. Имеется случайная выборка $(X_1, Y_1), ...(X_n, Y_n)$.

С помощью метода максимального правдоподобия оцените

1. θ и β;
2. $a = \theta/(\beta + \theta)$.

2)

Оценки коэффициентов линейной регрессии, полученные методом максимального правдоподобия и методом наименьших квадратов в случае нормально распределенной случайной составляющей, будут совпадать

- A всегда
- B никогда
- C если ковариационная матрица случайной составляющей нулевая
- D если ковариационная матрица случайной составляющей диагональна
- ■ если ковариационная матрица случайной составляющей пропорциональна единичной
- F *Нет верного ответа.*

3)

Оценки коэффициентов линейной регрессии, полученные методом максимального правдоподобия и методом наименьших квадратов в случае нормально распределенной случайной составляющей, будут совпадать

- A всегда
- B никогда
- C если ковариационная матрица случайной составляющей нулевая
- D если ковариационная матрица случайной составляющей диагональна
- ■ если ковариационная матрица случайной составляющей пропорциональна единичной
- F *Нет верного ответа.*

4)

Методом максимального правдоподобия Гоша оценил модель

$$Y_i = \beta_1 + \beta_2 X_{i2} + \ldots + \beta_6 X_{i6} + \varepsilon_i,$$

где $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2 I)$, по 12 наблюдениям. Оказалось, что $RSS = 24$. Оценка дисперсии случайной составляющей равна

A 0.5	C 2.4	■ 2
B 24/7	D 0.48	F *Нет верного ответа.*

5) Учебник Магнус, Катышев, Пересецкий

10.3. Пусть $y_1, \ldots, y_n$ — выборка из распределения с плотностью $h(y, \theta) = 1/\theta$, если $0 < x \leqslant \theta$, и $h(y, \theta) = 0$ — в остальных случаях $(0 < \theta < \infty)$. Покажите, что $\widehat{\theta} = \max y_i$ является оценкой максимального правдоподобия, и найдите ее смещение.

6) Учебник Магнус, Катышев, Пересецкий

10.6. Пусть $y_1, \ldots, y_n$ — независимые, одинаково распределенные случайные величины, равномерно распределенные на интервале $(\theta, 2\theta)$. Покажите, что:

а) оценка максимального правдоподобия есть $\widehat{\theta} = \max y_i/2$;

б) $\widehat{\theta}$ является смещенной, но асимптотически несмещенной;

в) $V(\widehat{\theta})$ асимптотически равна $\theta^2/(4n^2)$.

Задачи на проверку гипотез

7) Учебник Магнус, Катышев, Пересецкий

10.12. Пусть p — вероятность выпадения орла при бросании монеты. Из $n = 100$ испытаний $x = 42$ раза выпал орел и 58 — решка. Тестируйте на 5%-ном уровне значимости гипотезу $H_0 : p = 0.5$:

а) при помощи теста Вальда (W);

б) при помощи теста множителей Лагранжа (LM);

в) при помощи теста отношения правдоподобия (LR).

10.13. Имеется 80 наблюдений пуассоновской случайной величины X. Их среднее значение равно $\overline{x} = 1.7$. Тестируйте на 5%-ном уровне значимости гипотезу $H_0 : \lambda = 2.0$:

а) при помощи теста Вальда (W);

б) при помощи теста множителей Лагранжа (LM);

в) при помощи теста отношения правдоподобия (LR).

8) Из задачника Демешев-Борзых

Задача 5.1.

Дядя Вова (Владимир Николаевич) и Скрипач (Гедеван) зарабатывают на Плюке чатлы, чтобы купить гравицапу. Число заработанных за i-ый день чатлов имеет пуассоновское распределение, заработки за разные дни независимы. За прошедшие 100 дней они заработали 250 чатлов.

1. Оцените параметр λ пуассоновского распределения методом максимального правдоподобия.

Постройте 95% доверительный интервал для λ.

Проверьте гипотезу о том, что средний дневной заработок равен 2 чатла с помощью теста отношения правдоподобия, теста Вальда, теста множителей Лагранжа.

9) A.C.Cameron, P.K.Trivedy, Microeconometrics, p.235.

Дана выборка n независимых одинаково распределенных случайных величин $Y \sim N(\mu, 1)$. Проверьте гипотезу $H_0 : \mu = \mu^*$ с помощью теста отношения правдоподобия, теста Вальда и теста множителей Лагранжа.

10) Магнус, Катышев, Пересецкий, № 10.10

Известно, что в модели $Y = X\beta + \varepsilon$

Имеется гетероскедастичность, причем

$\mathrm{var}(\varepsilon_i) = \sigma_1^2,\ i = 1,\ldots,n_1,\ \ \mathrm{var}(\varepsilon_i) = \sigma_2^2,\ i = n_1+1,\ldots,n = n_1+n_2$

$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0,\ i \neq j$.

В предположении нормальности вектора ошибок постройте тест отношения правдоподобия (LR-test) для проверки гипотезы $H_0 : \sigma_1^2 = \sigma_2^2$.

11) Борзых, Демешев, № 5.19, 5.20

5.19 Исследователь Вениамин пытается понять, как логарифм количества решённых им по эконометрике задач зависит от количества съеденных им пирожков. Для этого он собрал 100 наблюдений. Первые 50 наблюдений — относятся к пирожкам с мясом, а последние 50 наблюдений — к пирожкам с повидлом. Вениамин считает, что ожидаемое количество решённых задач не зависит от начинки пирожков, а только от их количества, т.е. $y_i = \beta x_i + u_i$. Однако он полагает, что для пирожков с мясом — $u_i \sim \mathcal{N}(0;\sigma_M^2)$, а для пирожков с повидлом — $u_i \sim \mathcal{N}(0;\sigma_J^2)$.

1. Выпишите логарифмическую функцию правдоподобия.
2. Выпишите условия первого порядка для оценки $\beta, \sigma_M^2, \sigma_J^2$.

5.20 После долгих изысканий Вениамин пришёл к выводу, что $\beta = 0$, т.е. что логарифм количества решенных им по эконометрике за вечер задач имеет нормальное распределение y_i с математическим ожиданием ноль. Однако он по прежнему уверен, что дисперсия y_i зависит от того, какие пирожки он ел в этом вечер. Для пирожков с повидлом $y_i \sim \mathcal{N}(0;\sigma_J^2)$, а для пирожков с мясом — $y_i \sim \mathcal{N}(0;\sigma_M^2)$. Всего 100 наблюдений. Первые 50 вечеров относятся к пирожкам с мясом, последние 50 вечеров — к пирожкам с повидлом:

$$\sum_{i=1}^{50} y_i = 10,\ \sum_{i=1}^{50} y_i^2 = 100,\ \sum_{i=51}^{100} y_i = -10,\ \sum_{i=51}^{100} y_i^2 = 300$$

1. Найдите оценки σ_M^2, σ_J^2, которые получит Вениамин.

2. Помогите Вениамину проверить гипотезу $\sigma_M^2 = \sigma_J^2$ с помощью тестов отношения правдоподобия, множителей Лагранжа и Вальда.

Упражнение для работы с реальными данными

12) По данным файла rlms14.dta

1) Оцените зависимость заработной платы от возраста (квадратичная зависимость), пола, продолжительности рабочей недели в часах, количества подчиненных.
2) Проверьте, что максимальная заработная плата достигается в 40 лет с помощью

А) теста Вальда (использовав линейное и нелинейное ограничение),

Б) с помощью теста отношения правдоподобия.

3) Проверьте, что максимальная заработная плата достигается в 40 лет, а продолжительность рабочей недели не влияет на заработную плату с помощью

А) теста Вальда (использовав линейное и нелинейное ограничение),

Б) с помощью теста отношения правдоподобия.

Решение.

```
. gen age = 2005- birth_year

. reg wage age agesq gender duration_weekh subordinates

      Source |       SS       df       MS              Number of obs =    1014
-------------+------------------------------           F(  5,  1008) =   25.02
       Model |  6.7651e+09     5  1.3530e+09           Prob > F      =  0.0000
    Residual |  5.4510e+10  1008  54077771.4           R-squared     =  0.1104
-------------+------------------------------           Adj R-squared =  0.1060
       Total |  6.1275e+10  1013  60489099.1           Root MSE      =  7353.8

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   253.9315   133.9959     1.90   0.058    -9.011409    516.8743
       agesq |  -3.436652   1.557422    -2.21   0.028    -6.492812   -.3804918
      gender |  -3105.083   475.3481    -6.53   0.000    -4037.868   -2172.298
duration_weekh |   9.796951     20.996     0.47   0.641    -31.40392    50.99783
subordinates |   9.623287   1.218617     7.90   0.000     7.231969     12.0146
       _cons |   8279.367   2915.344     2.84   0.005     2558.529    14000.21
------------------------------------------------------------------------------

. est store reg1

. test age = -80*agesq

 ( 1)  age + 80*agesq = 0

       F(  1,  1008) =    0.91
            Prob > F =    0.3414

. testnl _b[age]/(2*_b[agesq])=-40

  (1)  _b[age]/(2*_b[agesq]) = -40

            F(1, 1008) =        0.61
              Prob > F =        0.4359
```

```
. gen x=-80*age+agesq

. reg wage x gender duration_weekh subordinates

      Source |       SS       df       MS              Number of obs =    1014
-------------+------------------------------           F(  4,  1009) =   31.05
       Model |  6.7161e+09     4  1.6790e+09           Prob > F      =  0.0000
    Residual |  5.4559e+10  1009  54072740.8           R-squared     =  0.1096
-------------+------------------------------           Adj R-squared =  0.1061
       Total |  6.1275e+10  1013  60489099.1           Root MSE      =  7353.4

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -3.961059   1.456666    -2.72   0.007    -6.819501   -1.102618
      gender |  -3124.049   474.9082    -6.58   0.000     -4055.97   -2192.129
duration_weekh |   11.30611   20.93508     0.54   0.589    -29.77518     52.3874
subordinates |   9.550631   1.216168     7.85   0.000     7.164122    11.93714
       _cons |   6605.916   2325.554     2.84   0.005      2042.44    11169.39
------------------------------------------------------------------------------

. est store reg2

. lrtest reg1 reg2

Likelihood-ratio test                                 LR chi2(1)  =      0.91
(Assumption: reg2 nested in reg1)                     Prob > chi2 =    0.3398

. reg wage age agesq gender duration_weekh subordinates

      Source |       SS       df       MS              Number of obs =    1014
-------------+------------------------------           F(  5,  1008) =   25.02
       Model |  6.7651e+09     5  1.3530e+09           Prob > F      =  0.0000
    Residual |  5.4510e+10  1008  54077771.4           R-squared     =  0.1104
-------------+------------------------------           Adj R-squared =  0.1060
       Total |  6.1275e+10  1013  60489099.1           Root MSE      =  7353.8

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   253.9315   133.9959     1.90   0.058    -9.011409    516.8743
       agesq |  -3.436652   1.557422    -2.21   0.028    -6.492812   -.3804918
      gender |  -3105.083   475.3481    -6.53   0.000    -4037.868   -2172.298
duration_weekh |   9.796951     20.996     0.47   0.641    -31.40392    50.99783
subordinates |   9.623287   1.218617     7.90   0.000     7.231969     12.0146
       _cons |   8279.367   2915.344     2.84   0.005     2558.529    14000.21
------------------------------------------------------------------------------

. est store reg3

. test (age = -80*agesq)  (duration_weekh=0)

 ( 1)  age + 80*agesq = 0
 ( 2)  duration_weekh = 0

       F(  2,  1008) =    0.60
            Prob > F =    0.5496

. testnl (_b[age]/(2*_b[agesq])=-40) (_b[duration_weekh]=0)

  (1)  _b[age]/(2*_b[agesq]) = -40
  (2)  _b[duration_weekh] = 0

            F(2, 1008) =        0.42
              Prob > F =        0.6567
```

```
. reg wage x gender  subordinates if duration_weekh !=.

      Source |       SS       df       MS              Number of obs =    1014
-------------+------------------------------           F(  3,  1010) =   41.33
       Model |  6.7003e+09     3  2.2334e+09           Prob > F      =  0.0000
    Residual |  5.4575e+10  1010  54034818.1           R-squared     =  0.1093
-------------+------------------------------           Adj R-squared =  0.1067
       Total |  6.1275e+10  1013  60489099.1           Root MSE      =  7350.8

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |  -4.079917   1.439439    -2.83   0.005     -6.90455   -1.255284
      gender |  -3178.134   464.0662    -6.85   0.000    -4088.778   -2267.489
subordinates |   9.582901   1.214273     7.89   0.000     7.200114    11.96569
       _cons |   7022.075   2193.404     3.20   0.001     2717.926    11326.23
------------------------------------------------------------------------------

. est store reg4

. lrtest reg3 reg4

Likelihood-ratio test                                 LR chi2(2)  =      1.20
(Assumption: reg4 nested in reg3)                     Prob > chi2 =    0.5477
```